Fundamentos matemáticos básicos de la realidad virtual

José Ignacio Ronda Prieto

Grupo de Tratamiento de Imágenes, ETSIT, UPM

jir@gti.ssr.upm.es

José Ignacio Ronda Prieto

jir@gti.ssr.upm.es

www.gti.ssr.upm.es/~jir

Grupo de Tratamiento de Imágenes

SSR - ETSIT - UPM

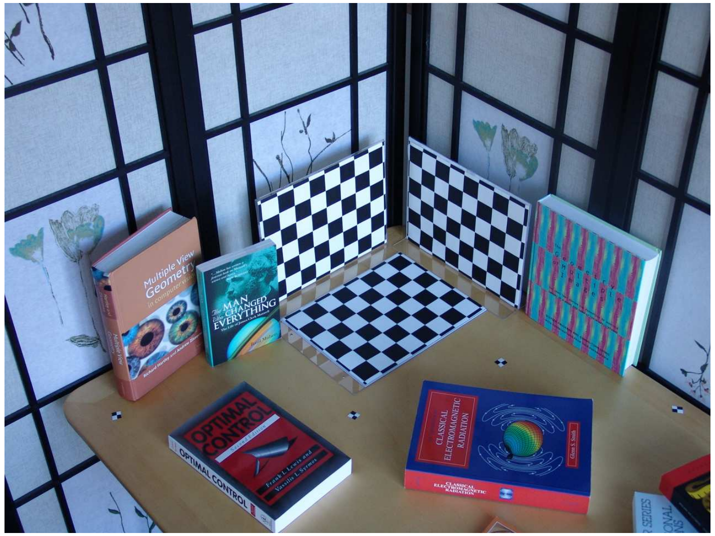
2012-2013

Cámaras y escenas 3D

En esta clase presentaremos las relaciones geométricas básicas entre los objetos 3D y sus imágenes.

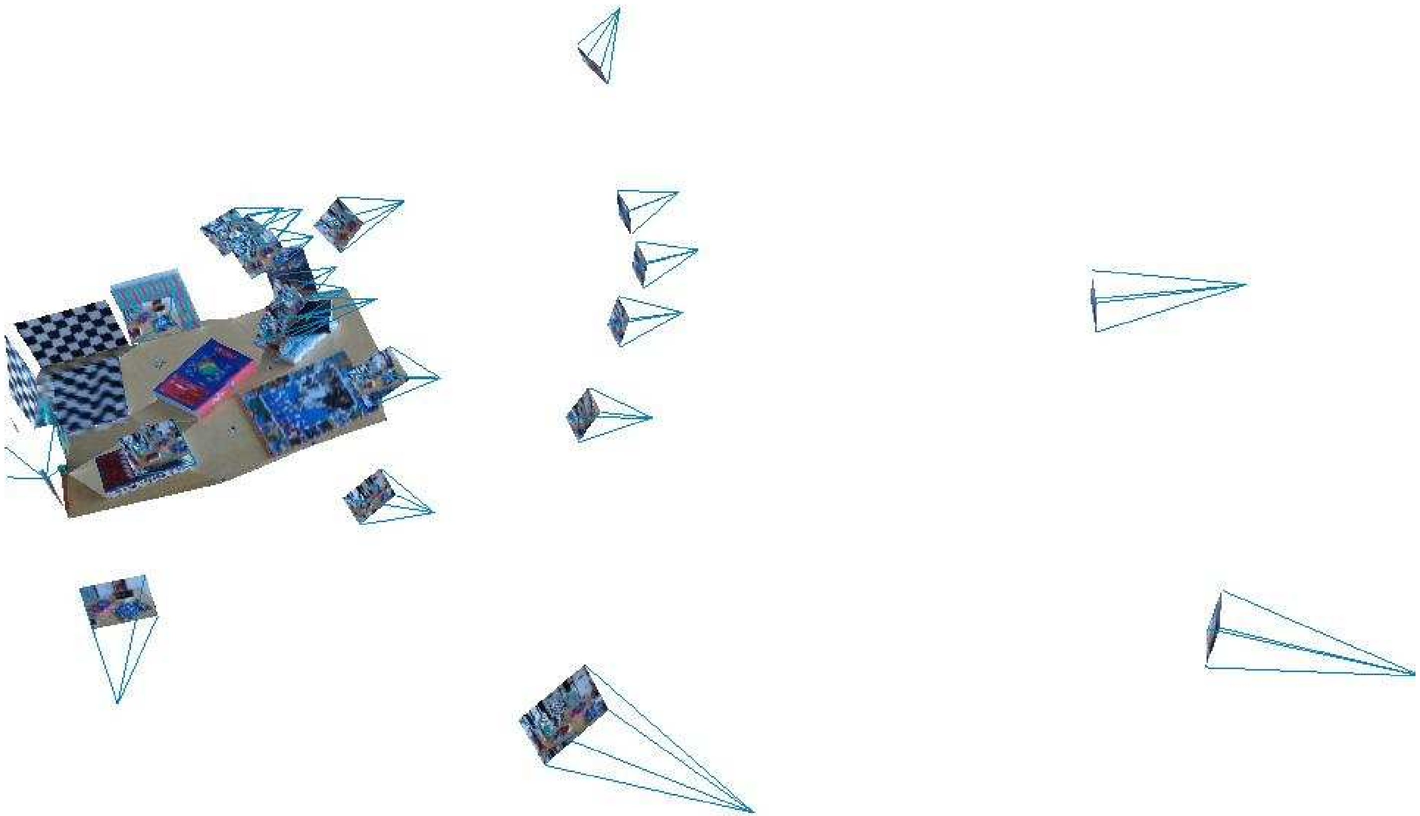
1. Introducción
2. Parámetros de la proyección
3. Calibración de la cámara
4. Coordenadas homogéneas
5. Relación epipolar
6. Aplicaciones

Aplicaciones: Reconstrucción 3D a partir de imágenes

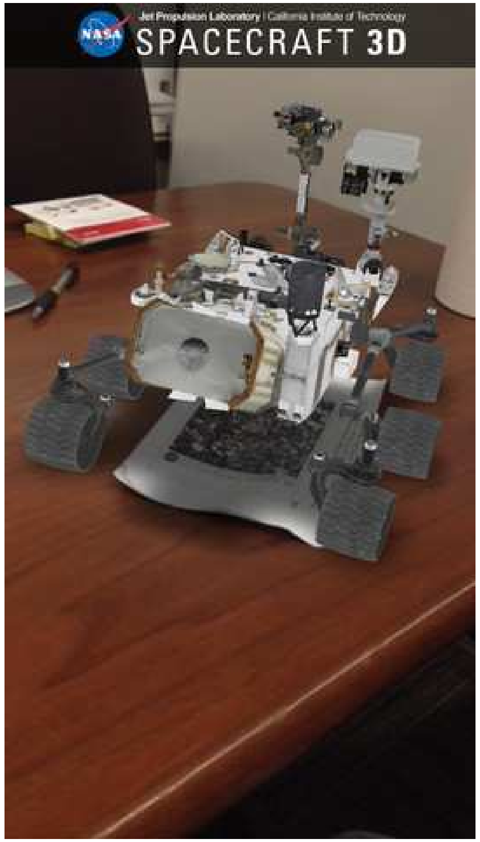


Aplicaciones: Reconstrucción 3D a partir de imágenes

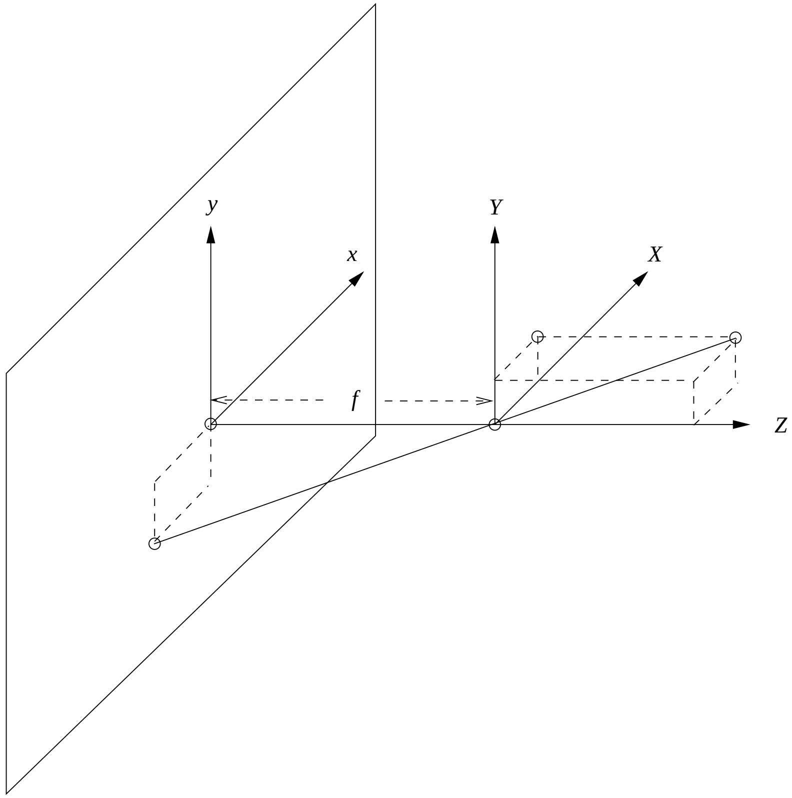
Aplicaciones: Reconstrucción 3D a partir de imágenes



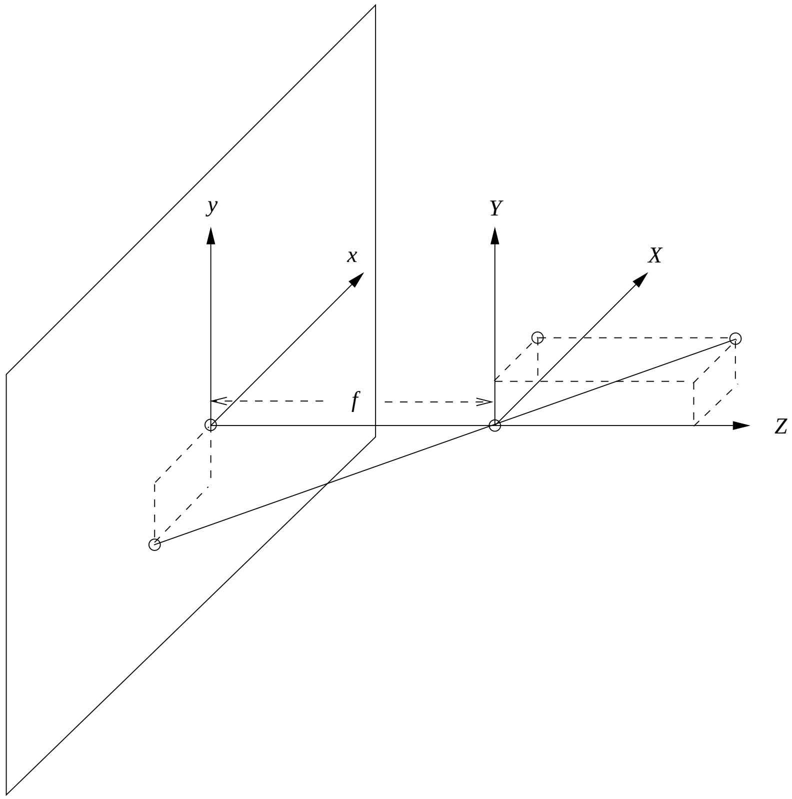
Aplicaciones: Realidad aumentada



Proyecciones

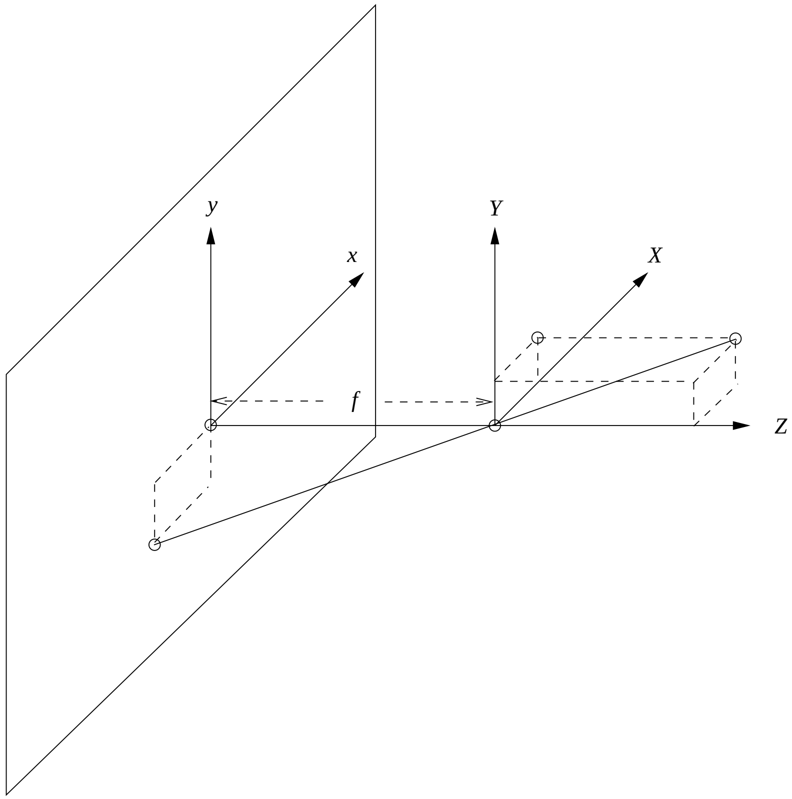


Proyecciones



Definen una correspondencia entre rectas que pasan por un punto *(centro de proyección)* y un plano *(plano de proyección)*. Los puntos del plano de proyección serán los puntos de la imagen.

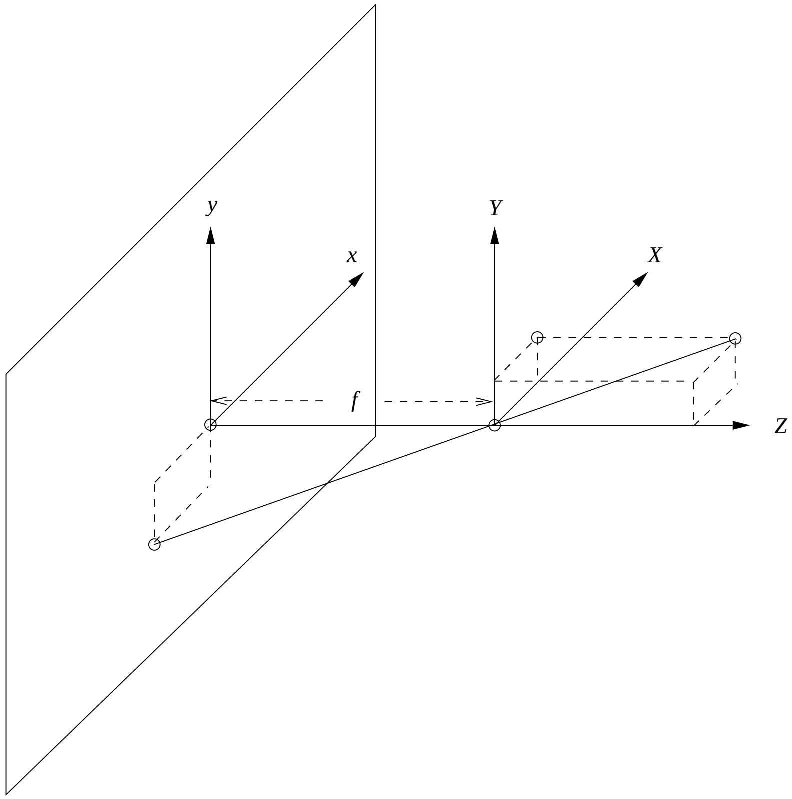
Proyecciones



Tomamos sistemas de referencia ortonormales:

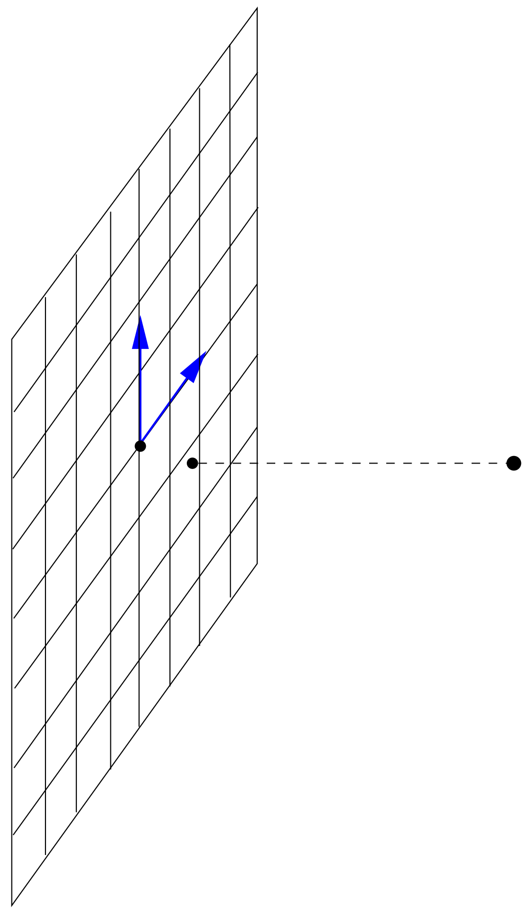
- El sistema de referencia en el espacio tiene origen en el centro de proyección y eje Z perpendicular al plano de proyección.
- El sistema de referencia (x, y) en el plano tiene origen en el punto más cercano al centro de proyección.
- Los ejes espaciales X e Y los tomamos paralelos a x e y.

Proyecciones



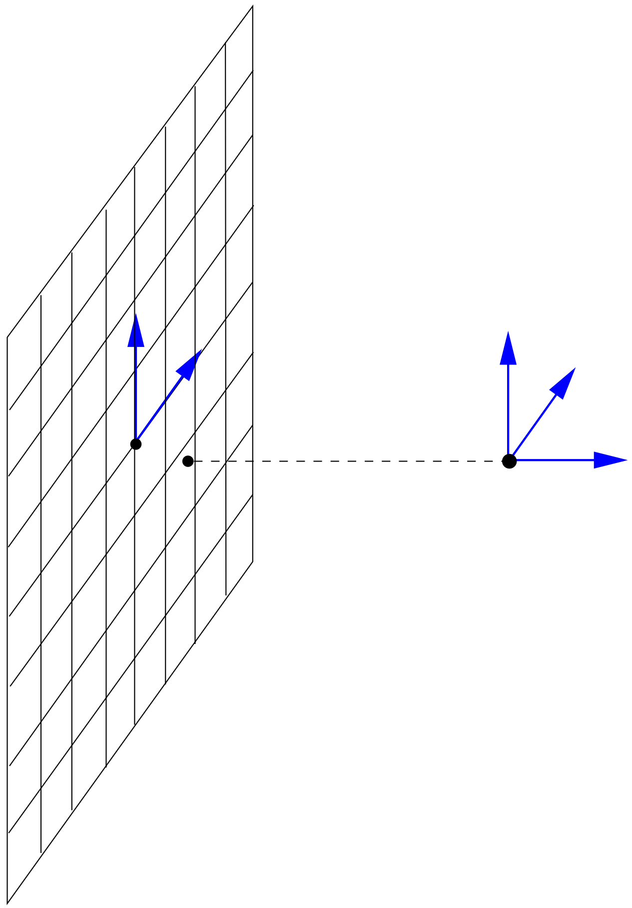
$$x = -f\frac{X}{Z}, \quad y = -f\frac{Y}{Z}$$

Parámetros intrínsecos



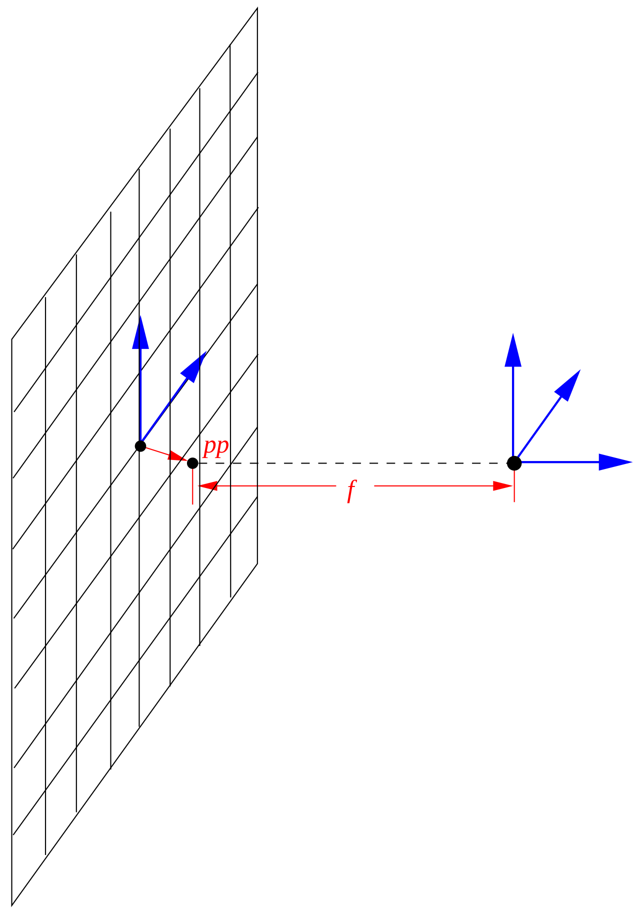
En la práctica no podemos usar exactamente estos sistemas de referencia porque no sabemos cuál es el punto del plano de proyección más cercano al centro de proyección.

Parámetros intrínsecos



El sistema de referencia (x, y) en la imagen tendrá un origen cualquiera y ejes ortonormales dados por los píxeles.
El sistema de referencia en el espacio será también ortonormal,tendrá origen en el centro de proyección y ejes (X, Y) paralelos a los (x, y).

Parámetros intrínsecos

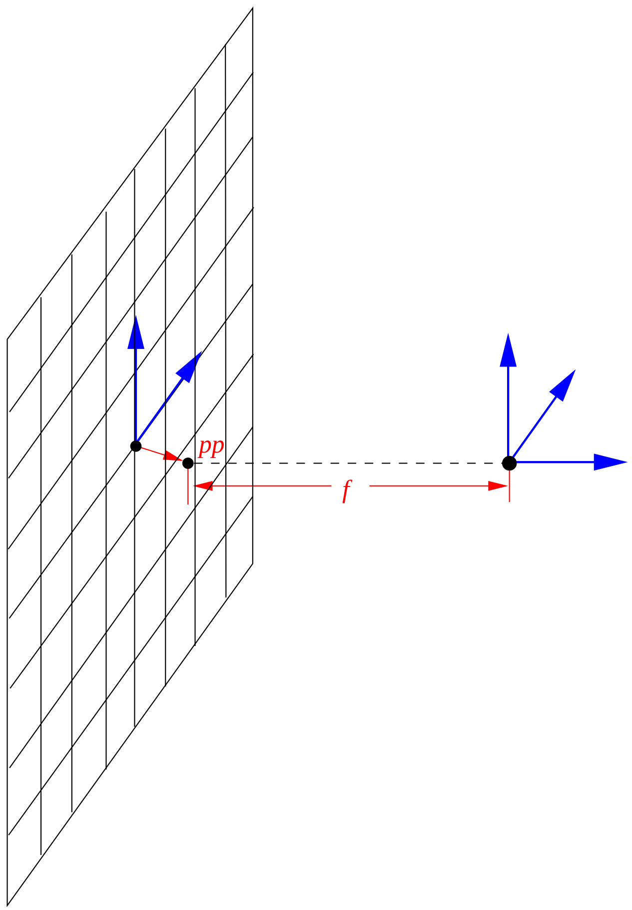


Parámetros intrínsecos:

Distancia focal (f)*:* Distancia entre el centro y el plano de proyección.

Punto principal (x_0, y_0)*:* Punto de la imagen más cercano al centro de proyección.

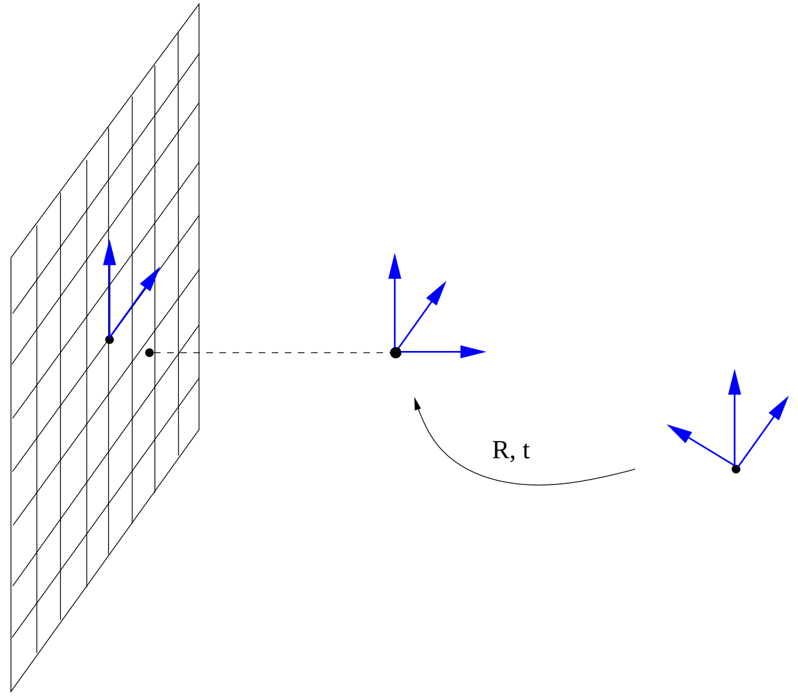
Parámetros intrínsecos



$$x = -f\frac{X}{Z} + x_0, \quad y = -f\frac{Y}{Z} + y_0$$

Parámetros extrínsecos

Los **parámetros extrínsecos** R, $\mathbf{t}$ indican la posición de la cámara en el espacio.

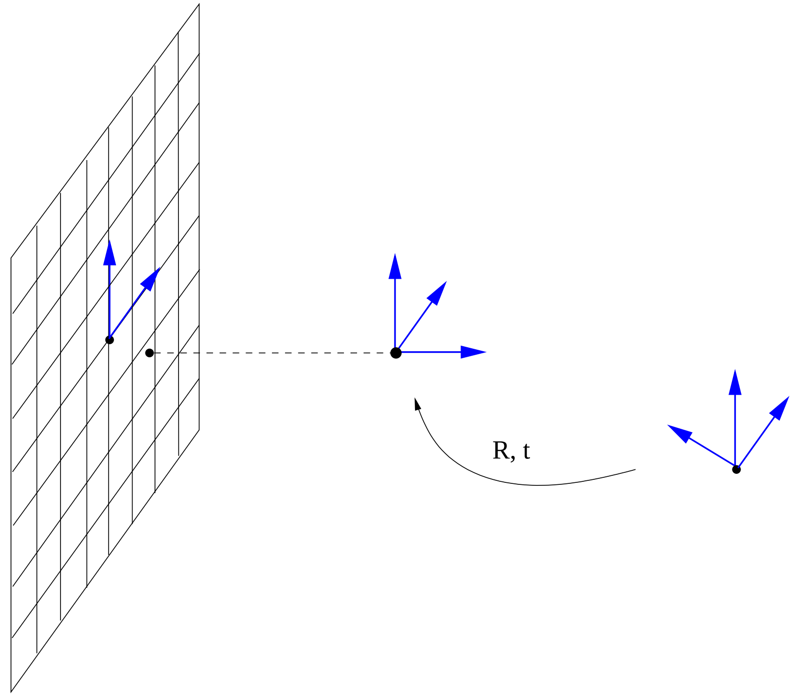


R es una matriz de rotación y $\mathbf{t}$ es un vector de traslación:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{\text{Cámara}} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{\text{Global}} + \mathbf{t}$$

Parámetros extrínsecos

Los **parámetros extrínsecos** R, $\mathbf{t}$ indican la posición de la cámara en el espacio.

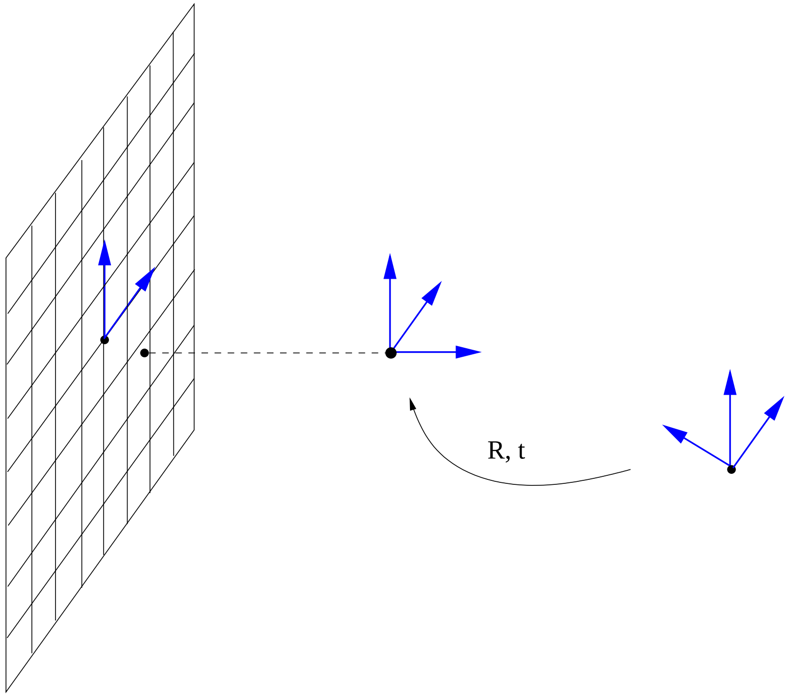


El centro de proyección tiene coordenadas $(0, 0, 0)$ en la referencia de la cámara, luego sus coordenadas globales $\mathbf{C}$ satisfacen

$$\mathbf{0} = R\mathbf{C} + \mathbf{t} \Rightarrow \mathbf{C} = -R^{-1}\mathbf{t} = -R^{\top}\mathbf{t}$$

Parámetros extrínsecos

Los **parámetros extrínsecos** R, $\mathbf{t}$ indican la posición de la cámara en el espacio.



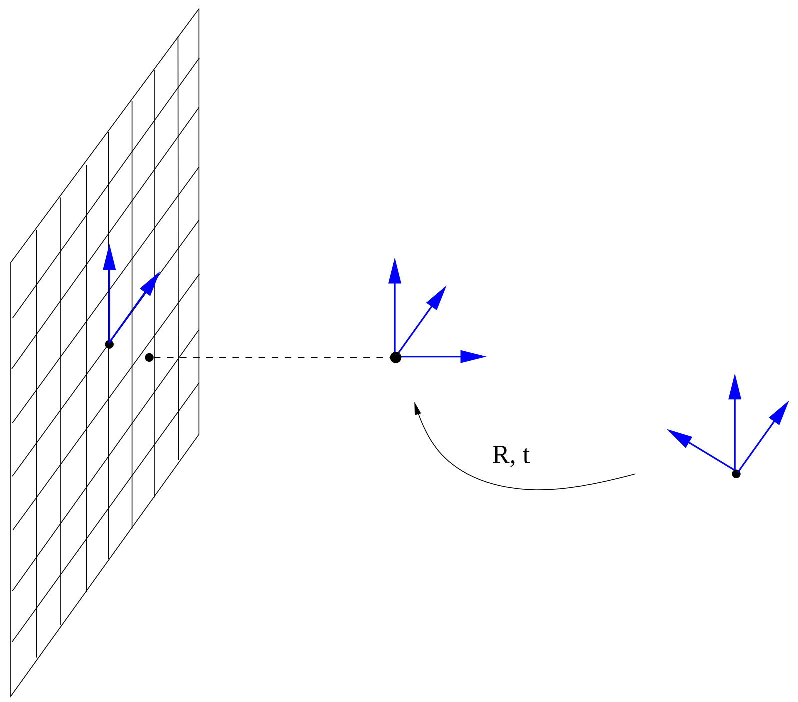
Si el punto con coordenadas (X, Y, Z) en la referencia global se ve en (x, y), tenemos las ecuaciones

$$x = -f\frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + x_0$$

$$y = -f\frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + y_0$$

Parámetros extrínsecos

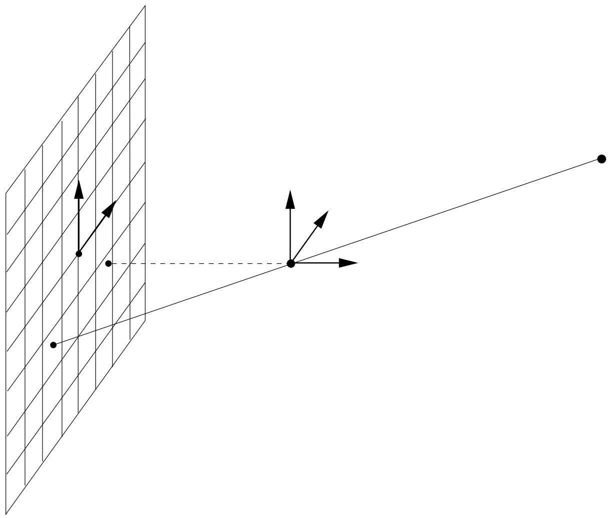
Los **parámetros extrínsecos** R, $\mathbf{t}$ indican la posición de la cámara en el espacio.

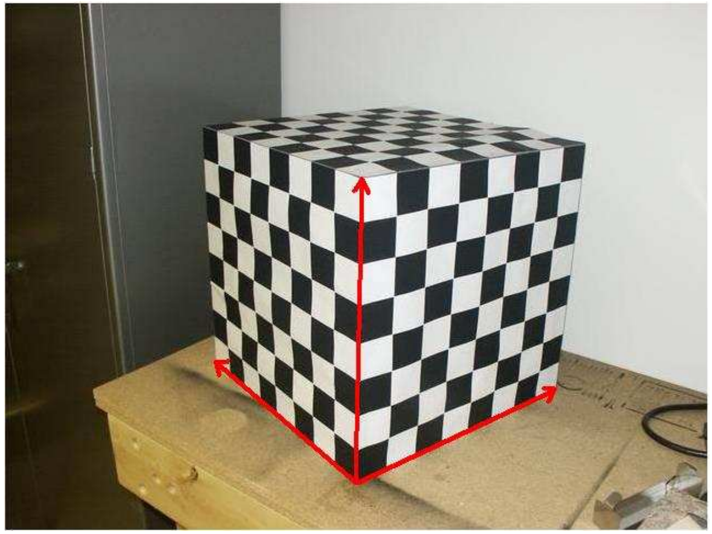


donde

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \\ \mathbf{r}_3^\top \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

Cálculo de los parámetros de la cámara





Conociendo las coordenadas de varios puntos 3D respecto de la referencia espacial y sus proyecciones podemos calcular los parámetros de la cámara.

Cálculo de los parámetros de la cámara

Si el punto con coordenadas (X, Y, Z) en la referencia global se ve en (x, y), tenemos las ecuaciones

$$x = -f\frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + x_0$$

$$y = -f\frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + y_0$$

donde

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \\ \mathbf{r}_3^\top \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

Cálculo de los parámetros de la cámara

Multiplicando por los denominadores obtenemos por cada punto dos ecuaciones

$$\begin{aligned} & xXr_{31} + xYr_{32} + xZr_{33} + xt_3 \\ =& X(x_0r_{31} + fr_{11}) + Y(x_0r_{32} + fr_{12}) + Z(x_0r_{33} + fr_{13}) + (x_0t_3 + ft_1) \\ & yXr_{31} + yYr_{32} + yZr_{33} + yt_3 \\ =& X(y_0r_{31} + fr_{21}) + Y(y_0r_{22} + fr_{12}) + Z(y_0r_{33} + fr_{23}) + (y_0t_3 + ft_2) \end{aligned}$$

que son lineales homogéneas si tomamos como variables

$$\begin{aligned} \mathbf{v}_1 &= x_0\mathbf{r}_3 + f\mathbf{r}_1 & \qquad w_1 &= x_0t_3 + ft_1 \\ \mathbf{v}_2 &= y_0\mathbf{r}_3 + f\mathbf{r}_2 & w_2 &= y_0t_3 + ft_2 \\ \mathbf{v}_3 &= \mathbf{r}_3 & w_3 &= t_3 \end{aligned}$$

Cálculo de los parámetros de la cámara

Como el sistema es homogéneo, obtendremos la solución salvo por un factor de escala común α:

$$\begin{aligned} \mathbf{v}_1 &= \alpha(x_0\mathbf{r}_3 + f\mathbf{r}_1) & w_1 &= \alpha(x_0 t_3 + f t_1) \\ \mathbf{v}_2 &= \alpha(y_0\mathbf{r}_3 + f\mathbf{r}_2) & w_2 &= \alpha(y_0 t_3 + f t_2) \\ \mathbf{v}_3 &= \alpha\mathbf{r}_3 & w_3 &= \alpha t_3 \end{aligned}$$

Para determinar estos $3 \times 3 + 3 = 12$ datos habremos necesitado al menos 6 puntos ($6 \times 2 = 12$).

De las variables obtenemos los parámetros utilizando que

- $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ es un sistema ortonormal.
- $\det(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = 1$.
- $f > 0$.

Cálculo de los parámetros de la cámara

Para ello procedemos como indicamos a continuación.
Suponemos escalados todos los datos de forma que $\|\mathbf{v}_3\| = 1$, luego $\alpha = \pm 1$.

1. Tenemos

$$\det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \alpha^3 \det(x_0\mathbf{r}_3 - f\mathbf{r}_1, y_0\mathbf{r}_3 - f\mathbf{r}_2, \mathbf{r}_3)$$
$$= \alpha^3 \det(-f\mathbf{r}_1, -f\mathbf{r}_2, \mathbf{r}_3) = \alpha^3 f^2 \det(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \alpha^3 f^2 = \alpha f^2$$

   luego el signo de $\det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ nos da α y la raíz cuadrada de su valor absoluto nos proporciona f.

2. Vamos despejando sucesivamente

$$\begin{aligned}
\mathbf{r}_3 &= \mathbf{v}_3/\alpha, & & \\
x_0 &= \langle \mathbf{v}_1/\alpha, \mathbf{r}_3 \rangle, & y_0 &= \langle \mathbf{v}_2/\alpha, \mathbf{r}_3 \rangle, \\
\mathbf{r}_1 &= \left(\frac{\mathbf{v}_1}{\alpha} - x_0\mathbf{r}_3 \right) / f, & \mathbf{r}_2 &= \left(\frac{\mathbf{v}_2}{\alpha} - y_0\mathbf{r}_3 \right) / f, \\
t_3 &= w_3/\alpha & & \\
t_1 &= (w_1/\alpha - x_0 t_3)/f, & t_2 &= (w_2/\alpha - x_0 t_3)/f
\end{aligned}$$

Puntos de fuga

Puntos de fuga

Suponemos ahora que las coordenadas de los puntos 3D están dadas en la referencia asociada a la cámara.
Cada punto de la recta $(X, Y, Z) = (X_0, Y_0, Z_0) + \lambda(v_1, v_2, v_3)$ proyectará sobre

$$x(\lambda) = x_0 - f \frac{X_0 + \lambda v_1}{Z_0 + \lambda v_3}$$

$$y(\lambda) = y_0 - f \frac{Y_0 + \lambda v_2}{Z_0 + \lambda v_3}$$

Los puntos para λ muy grande se acercarán a

$$x_\infty = x_0 - f \frac{v_1}{v_3}, \; y_\infty = y_0 - f \frac{v_2}{v_3}$$

Puntos de fuga

$$x_\infty = x_0 - f\frac{v_1}{v_3},\ y_\infty = y_0 - f\frac{v_2}{v_3}$$

Como este punto no depende de (X_0, Y_0, Z_0), resulta que es el punto en el que se cortan las proyecciones de todas las rectas con el mismo vector director $\mathbf{v}$.
Decimos que (x_∞, y_∞) es el *punto de fuga* de la recta.

Puntos de fuga

$$x_\infty = x_0 - f\frac{v_1}{v_3},\ y_\infty = y_0 - f\frac{v_2}{v_3}$$

Si conozco los parámetros intrínsecos, de (x_∞, y_∞) puedo obtener la dirección $\mathbf{v} = (v_1, v_2, v_3)$:

$$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \sim \begin{pmatrix} (x_0 - x_\infty)/f & (y_0 - y_\infty)/f & 1 \end{pmatrix}$$
$$\sim \begin{pmatrix} x_0 - x_\infty & y_0 - y_\infty & f \end{pmatrix}$$

Calibración con puntos de fuga

Un algoritmo muy utilizado se basa en imágenes de un tablero de ajedrez.
Una imagen como ésta proporciona dos pares de puntos de fuga correspondientes a direcciones ortogonales.

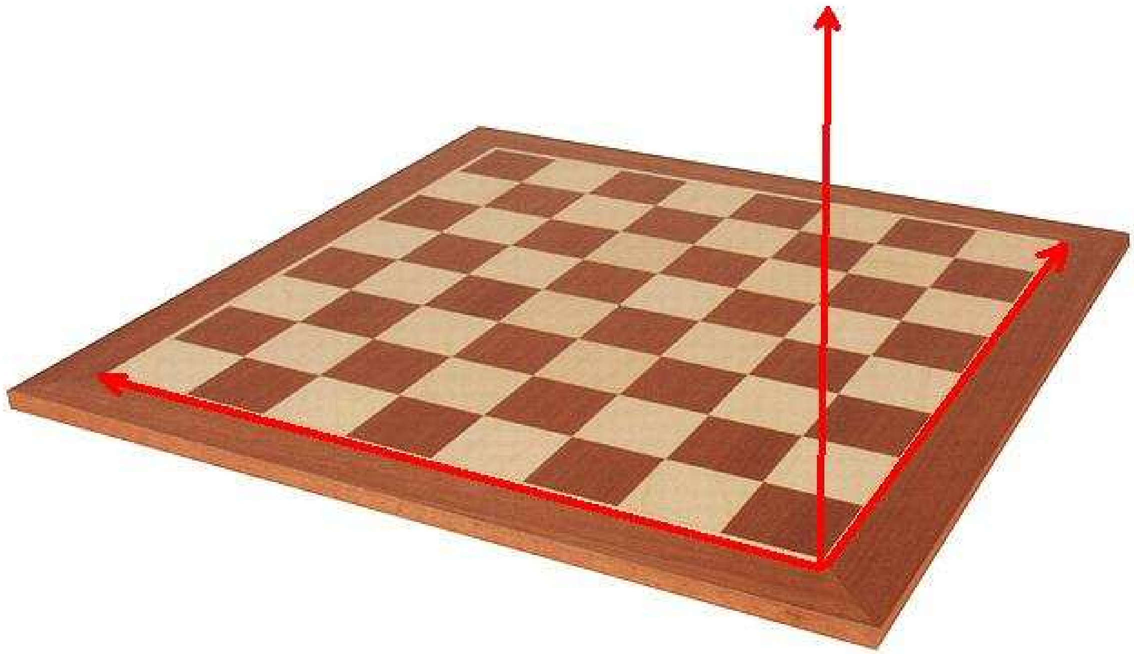
Calibración con puntos de fuga

Si $(x_\infty^{(1)}, y_\infty^{(1)})$ y $(x_\infty^{(2)}, y_\infty^{(2)})$ son proyecciones de dos direcciones ortogonales $\mathbf{v}_1$, $\mathbf{v}_2$ tenemos la ecuación

$$\begin{aligned}\langle \mathbf{v}_1, \mathbf{v}_2 \rangle &= (x_0 - x_\infty^{(1)})(x_0 - x_\infty^{(2)}) + (y_0 - y_\infty^{(1)})(y_0 - y_\infty^{(2)}) + f^2 \\ &= -x_0(x_\infty^{(1)} + x_\infty^{(2)}) + x_\infty^{(1)} x_\infty^{(2)} - y_0(y_\infty^{(1)} + y_\infty^{(2)}) + y_\infty^{(1)} y_\infty^{(2)} \\ &+ x_0^2 + y_0^2 + f^2 = 0,\end{aligned}$$

que es una ecuación lineal no homogénea en las incógnitas $x_0, y_0, (x_0^2 + y_0^2 + f^2)$.
Por tanto las imágenes de dos tableros (cuatro ecuaciones) bastan para obtener los parámetros intrínsecos $(x_0,\ y_0,\ f)$ de la cámara.

Calibración con puntos de fuga



Una vez que conocemos los parámetros intrínsecos, es fácil obtener la posición de la cámara respecto de la referencia dada por el tablero (parámetros extrínsecos).

Calibración con puntos de fuga

Cálculo de los parámetros extrínsecos:

- Las cuatro esquinas del tablero son los puntos 3D de coordenadas $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(1,1,0)$.
- Conociendo sus proyecciones y los parámetros intrínsecos podemos obtener los vectores $\mathbf{s}_1$, $\mathbf{s}_2$ (columnas de la matriz R) y $\mathbf{t}$ salvo por un factor de escala común.
- De ahí calculamos estos tres vectores, utilizando que $\|\mathbf{s}_1\| = \|\mathbf{s}_2\| = 1$.
- Calculamos $\mathbf{s}_3 = \mathbf{s}_1 \times \mathbf{s}_2$.

Autocalibracion

¿Podemos calibrar sin saber nada de la escena?

Autocalibracion

La *autocalibración* permite obtener los parámetros de las cámaras a partir de correspondencias de puntos con posiciones 3D completamente desconocidas.

Para que sea matemáticamente posible las cámaras deben verificar alguna restricción, como

- Tener todas los mismos parámetros.
- O simplemente tener píxeles cuadrados.

Más información en

www.gti.ssr.upm.es/˜jir/comp_vis

Coordenadas homogéneas en el plano

Ecuación *afín* de una recta:

$$ax + by + c = 0$$

Coordenadas homogéneas de la recta:

$$\mathbf{r} \sim (a, b, c)$$

Punto de la recta

$$(x, y), \quad (x, y, 1)(a, b, c)^\top = 0$$

Coordenadas homogéneas del punto:

$$(X, Y, T) \sim (x, y, 1), \qquad \frac{X}{T} = x, \ \frac{Y}{T} = y$$

Ecuación de la recta en coordenadas homogéneas

$$ax + by + c = 0 \Leftrightarrow a\frac{X}{T} + b\frac{Y}{T} + c = 0 \Leftrightarrow aX + bY + cT = 0$$

Coordenadas homogéneas en el plano

Punto de corte de dos rectas

$$aX + bY + cT = 0, \quad a'X + b'Y + c'T = 0$$

$$(X, Y, T) \perp (a, b, c), \; (a', b', c')$$

$$\Rightarrow (X, Y, T) \sim (a, b, c) \times (a', b', c').$$

Recta por dos puntos

$$aX + bY + cT = 0, \quad aX' + bY' + cT' = 0$$

$$(a, b, c) \perp (X, Y, T), \; (X', Y', T')$$

$$\Rightarrow (a, b, c) \sim (X, Y, T) \times (X', Y', T').$$

Coordenadas homogéneas en el plano

Punto de corte de dos rectas

$$aX + bY + cT = 0, \quad a'X + b'Y + c'T = 0$$

$$(X, Y, T) \perp (a, b, c), \ (a', b', c')$$

$$\Rightarrow (X, Y, T) \sim (a, b, c) \times (a', b', c').$$

Recta por dos puntos

$$aX + bY + cT = 0, \quad aX' + bY' + cT' = 0$$

$$(a, b, c) \perp (X, Y, T), \ (X', Y', T')$$

$$\Rightarrow (a, b, c) \sim (X, Y, T) \times (X', Y', T').$$

Cortamos dos rectas *paralelas*

$$aX + bY + cT = 0, \quad aX + bY + c'T = 0$$

$$\Rightarrow (X, Y, T) \sim (a, b, c) \times (a, b, c') \sim (b, -a, 0)$$

Obtenemos el *vector director* de la recta.

Coordenadas homogéneas en el plano

$(-b, a, 0)$ es el “punto” en el que se cortan todas las rectas con vector director $\mathbf{v} = (-b, a)$

$\Rightarrow$ *punto del infinito* de la recta.

Los puntos del infinito forman la *recta del infinito,* $T = 0$.

Coordenadas homogéneas en el plano

$(-b, a, 0)$ es el “punto” en el que se cortan todas las rectas con vector director $\mathbf{v} = (-b, a)$
$\Rightarrow$ *punto del infinito* de la recta.
Los puntos del infinito forman la *recta del infinito,* $T = 0$.

Coordenadas homogéneas en el espacio

Análogamente:

- Las **coordenadas homogéneas de un punto** (x, y, z) del espacio 3D son $(x, y, z, 1)$ o cualquier vector no nulo proporcional $(X, Y, Z, T) = \alpha(x, y, z, 1)$.
- Las **coordenadas homogéneas de un plano** de ecuación $ax + by + cz + d = 0$ son (a, b, c, d) o cualquier vector no nulo proporcional.
- Los vectores de la forma $(U, V, W, 0)$ representan coordenadas homogéneas de **puntos del infinito.**
- El plano de ecuación $T = 0$ es el **plano del infinito**.

Coordenadas homogéneas en el espacio

Utilizando coordenadas homogéneas en el espacio y en la imagen, la ecuación de la proyección se puede expresar de la forma

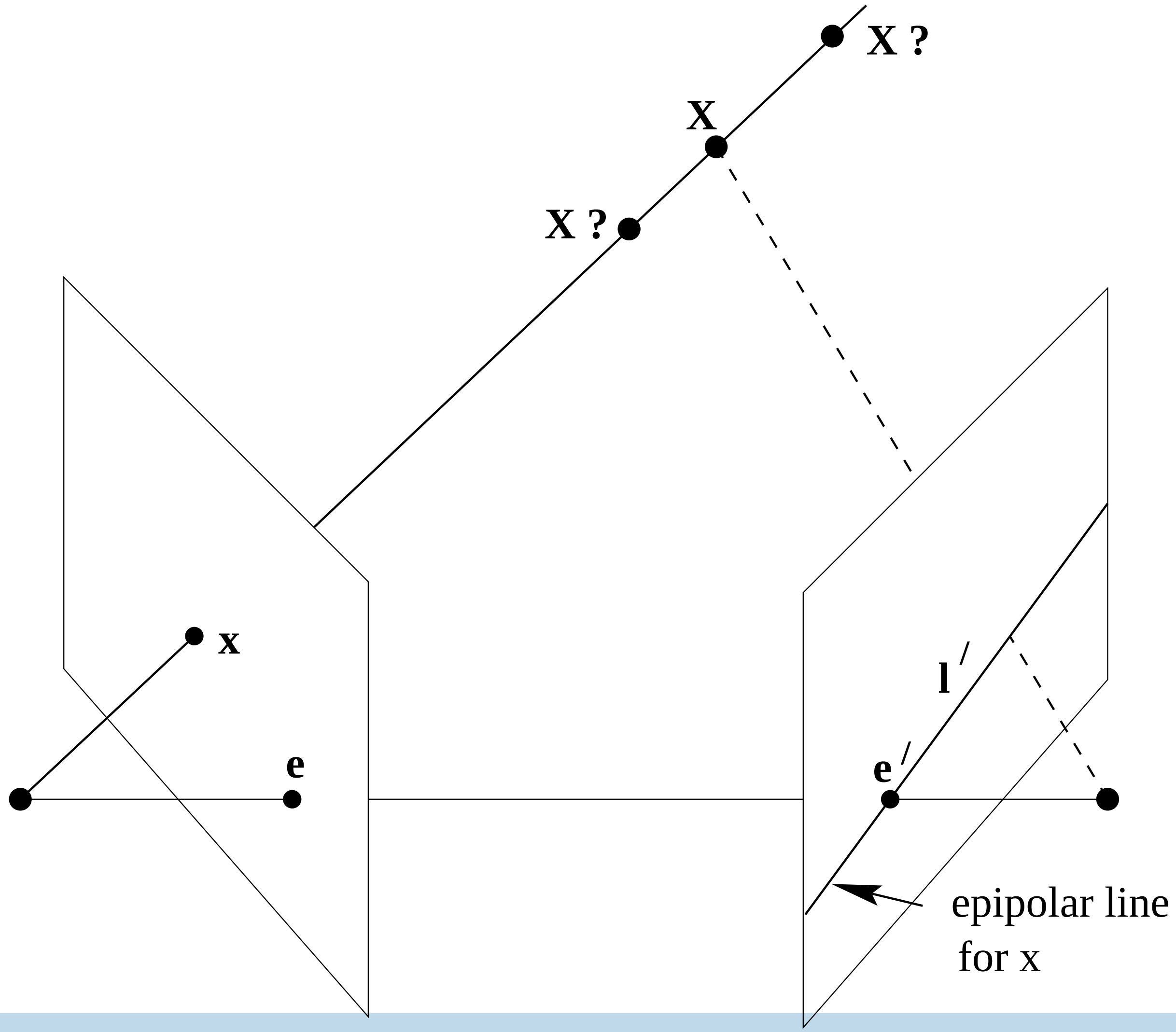
$$\begin{pmatrix} x \\ y \\ t \end{pmatrix} = \underbrace{\underbrace{\begin{pmatrix} -f & & x_0 \\ & -f & y_0 \\ & & 1 \end{pmatrix}}_{K} \begin{pmatrix} R & \mathbf{t} \end{pmatrix}}_{P} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

El centro de proyección es el núcleo de P:

$$P \begin{pmatrix} \mathbf{C} \\ 1 \end{pmatrix} = K \underbrace{\begin{pmatrix} R & \mathbf{t} \end{pmatrix} \begin{pmatrix} \mathbf{C} \\ 1 \end{pmatrix}}_{R\mathbf{C} + \mathbf{t} = \mathbf{0}} = \mathbf{0}$$

La relación epipolar

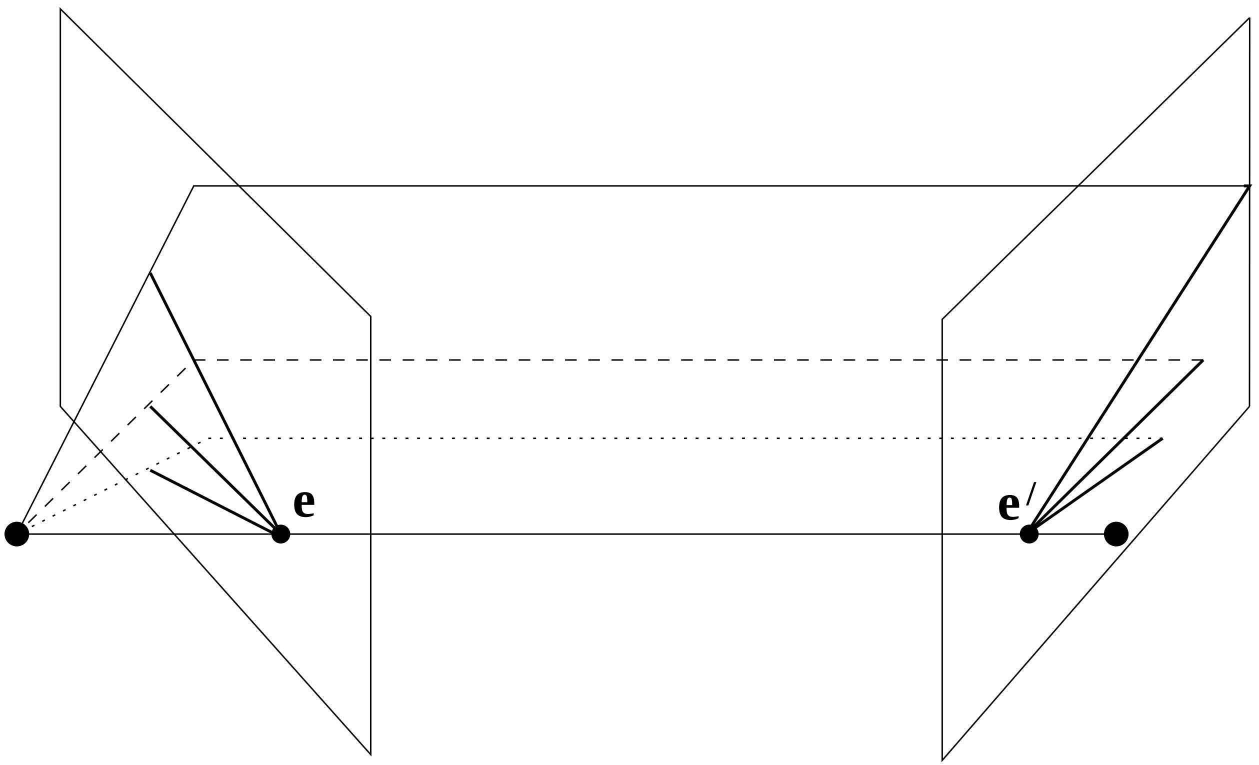
Si veo un punto proyectado en una imagen, ¿dónde puede estar su proyección en la otra?



La relación epipolar

Puntos correspondientes están sobre rectas correspondientes de los haces con bases en los *epipolos* e, e', que son las imágenes de cada centro de proyección $\mathbf{C}$, $\mathbf{C}'$, en la otra imagen:

$$\mathbf{e} \sim PC', \quad \mathbf{e}' \sim PC.$$



La relación epipolar

Ecuación de la relación epipolar

Usamos coordenadas homogéneas.
Si $\mathbf{q}$ y $\mathbf{q}'$ son las proyecciones del punto $\mathbf{Q}$,

$$\alpha\mathbf{q} = P\mathbf{Q}$$

$$\alpha'\mathbf{q}' = P'\mathbf{Q}$$

Estas ecuaciones se pueden escribir de la siguiente forma

$$\begin{pmatrix} P & \mathbf{q} & \\ P' & & \mathbf{q}' \end{pmatrix} \begin{pmatrix} \mathbf{Q} \\ -\alpha \\ -\alpha' \end{pmatrix} = \mathbf{0}$$

Esta ecuación indica que la matriz es singular, por tanto

$$\det \begin{pmatrix} P & \mathbf{q} & \\ P' & & \mathbf{q}' \end{pmatrix} = 0$$

La relación epipolar

Este determinante es una expresión lineal en $\mathbf{q}$ y $\mathbf{q}'$, por lo que debe poder escribirse de la forma

$$\det\begin{pmatrix} P & \mathbf{q} & \\ P' & & \mathbf{q}' \end{pmatrix} = \mathbf{q}'^{\top} F \mathbf{q} = 0.$$

Interpretación:

- $F\mathbf{q}$ son las coordenadas de la recta de los puntos en los que puede estar el correspondiente a $\mathbf{q}$.
- $F^{\top}\mathbf{q}$ idem para $\mathbf{q}'$.

La relación epipolar

La matriz F se denomina *matriz fundamental*. Es una matriz **singular:** En efecto, dado un $\mathbf{q}$ cualquiera,

$$\mathbf{e}'^{\top} F \mathbf{q} = \det \begin{pmatrix} P & \mathbf{q} & \\ P' & & \mathbf{e}' \end{pmatrix} = \det \begin{pmatrix} P & \mathbf{q} & P\mathbf{C} \\ P' & & P'\mathbf{C} \end{pmatrix}$$

Este determinante es cero porque la última fila es combinación lineal de las cuatro primeras con coeficientes dados por las coordenadas del vector $\mathbf{C}$.
Como esto es así para todo $\mathbf{q}$ tenemos que $\mathbf{e}'^{\top} F = \mathbf{0}$, luego F es singular y su núcleo por la izquierda corresponde a $\mathbf{e}'$.
Análogamente se comprueba que su núcleo por la derecha corresponde a $\mathbf{e}$.

La relación epipolar

En resumen: el punto (x, y) de una imagen y el punto (x', y') de la otra pueden corresponder a un mismo punto 3D si

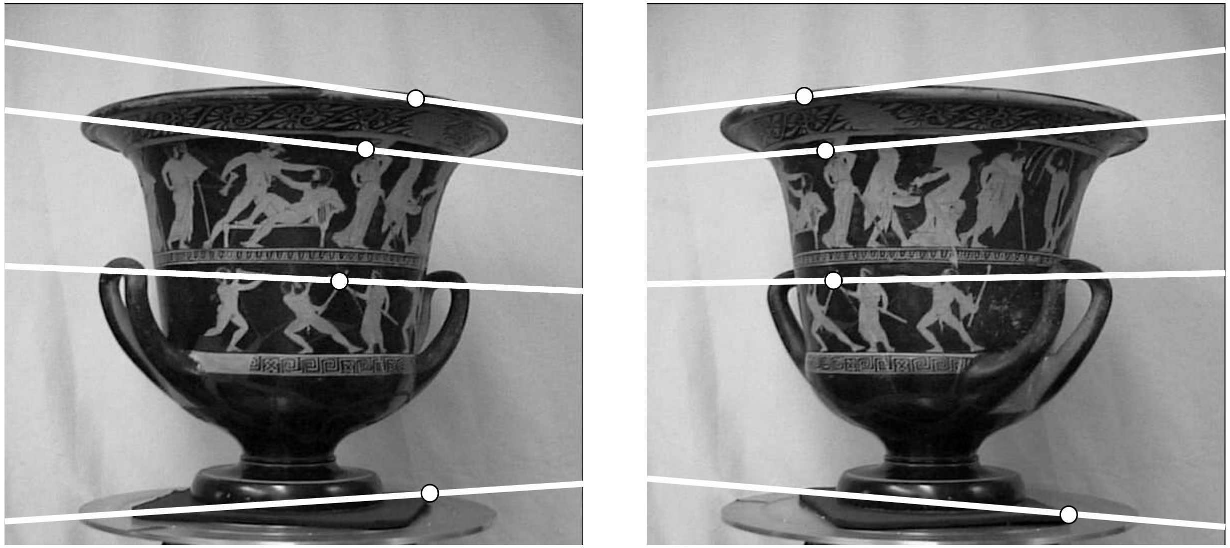
$$\begin{pmatrix} x' & y' & 1 \end{pmatrix} F \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$$

(condición necesaria, no suficiente)
La matriz fundamental se puede calcular a partir de ocho correspondencias

$$(x_i, y_i) \leftrightarrow (x'_i, y'_i)$$

despejando los coeficientes de F del sistema lineal dado por las ecuaciones.

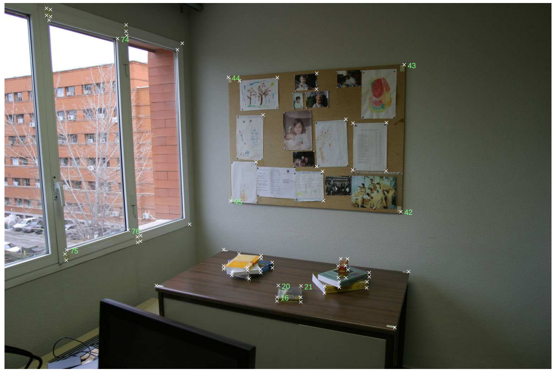
La relación epipolar

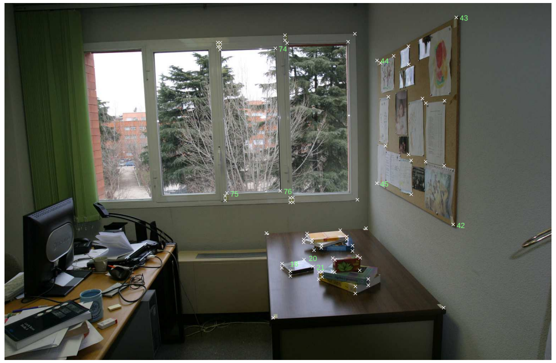


Una vez conocida la *relación epipolar* es más fácil buscar más correspondencias entre puntos de dos imágenes.

Aplicaciones: Reconstrucción 3D

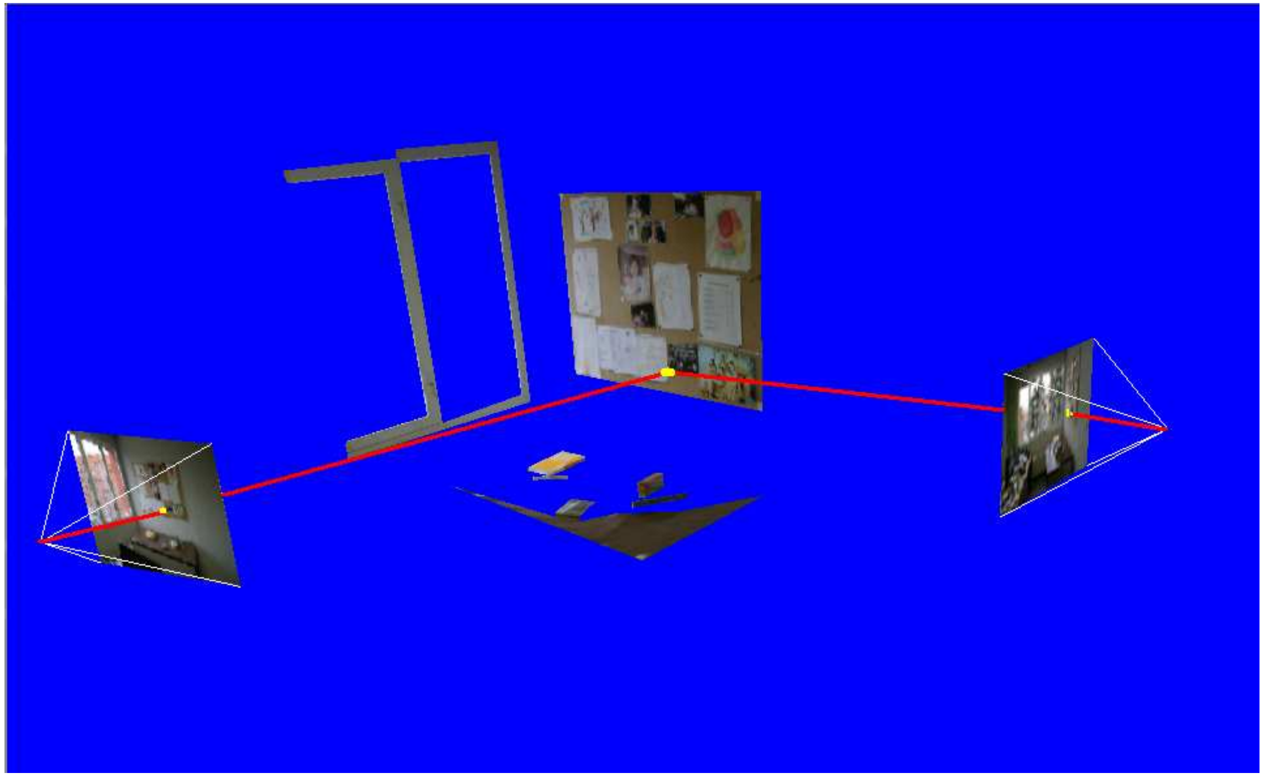
1. Calibramos las cámaras.
2. Detectamos varias correspondencias fiables y calculamos la matriz fundamental.
3. Utilizamos la matriz fundamental para detectar más correspondencias.





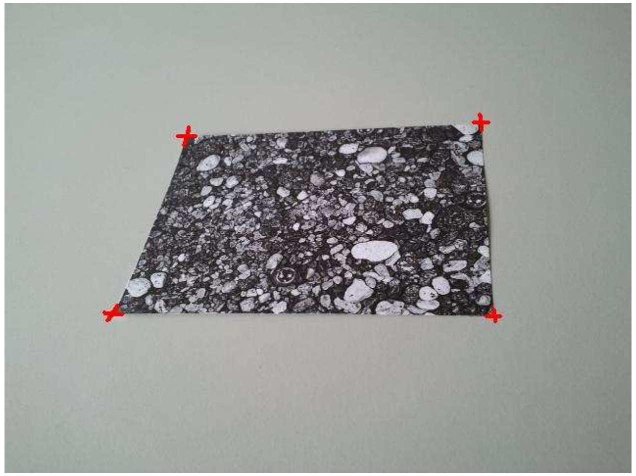
Aplicaciones: Reconstrucción 3D

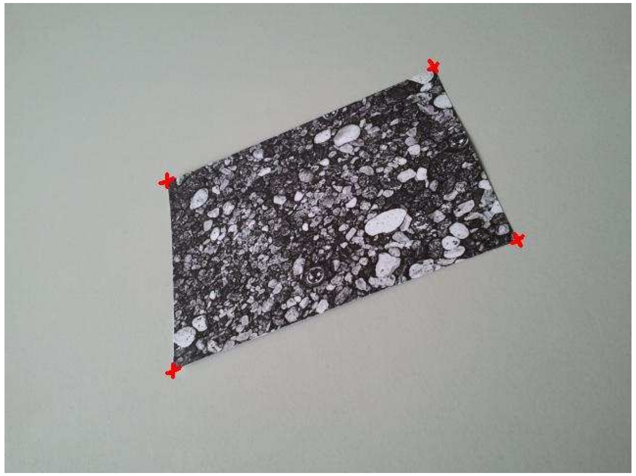
4. Resolvemos las ecuaciones de la proyección para cada correspondencia.



Aplicaciones: Realidad aumentada

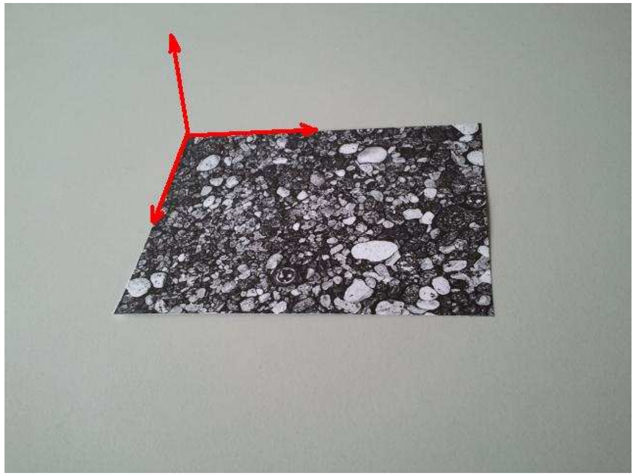
1. Detectamos en varias imágenes las esquinas del marcador.
2. Obtenemos los parámetros intrínsecos de la cámara con el algoritmo de calibración basado en puntos de fuga.

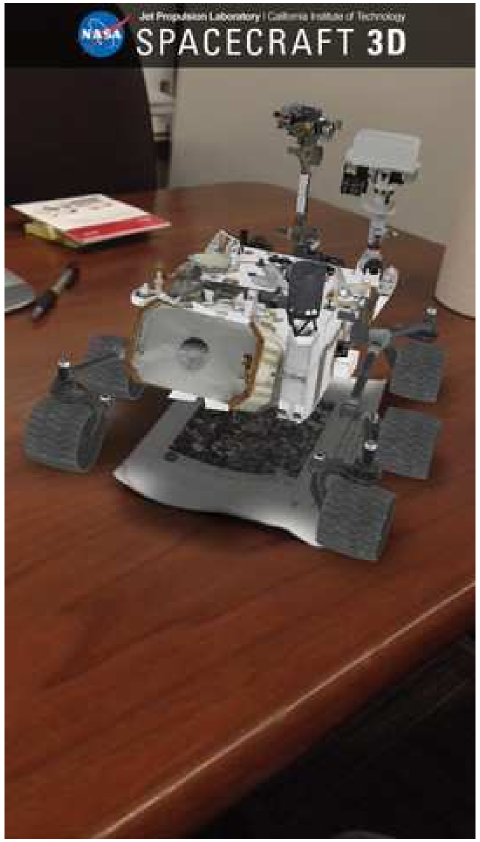




Aplicaciones: Realidad aumentada

3. Para cada nueva imagen
3.1. Detectamos las esquinas del marcador.
3.2. Obtenemos los parámetros extrínsecos de la cámara.
3.3. Representamos la proyección del objeto.





Bibliografía

- R. Hartley, A. Zisserman, “Multiple View Geometry in Computer Vision”, Cambridge Univ. Press, 2^a ed., Reino Unido, 2003.